確率・統計基礎: 点推定、不偏推定量

森 立平

二乗誤差

$$
\begin{aligned}
\mathrm{MSE}(\theta, \delta) &:= \mathbb{E}\left[(\delta(X) - g(\theta))^2 \mid \theta\right] \\
&= \mathbb{E}\left[\left(\delta(X) \color{red}{- \mathbb{E}[\delta(X) \mid \theta] + \mathbb{E}[\delta(X) \mid \theta]} - g(\theta)\right)^2 \mid \theta\right] \\
&= \mathbb{E}\left[(\delta(X) - \mathbb{E}[\delta(X) \mid \theta])^2 \mid \theta\right] + \mathbb{E}\left[(\mathbb{E}[\delta(X) \mid \theta] - g(\theta))^2 \mid \theta\right] \\
&\quad + 2\mathbb{E}\left[(\delta(X) - \mathbb{E}[\delta(X) \mid \theta])(\mathbb{E}[\delta(X) \mid \theta] - g(\theta)) \mid \theta\right] \\
&= \mathbb{V}[\delta(X) \mid \theta] + (\mathbb{E}[\delta(X) \mid \theta] - g(\theta))^2 \\
&= \mathbb{V}[\delta(X) \mid \theta] + \mathrm{Bias}_g(\delta \mid \theta)^2
\end{aligned}
$$

$$
\mathrm{Bias}_g(\delta \mid \theta) := \mathbb{E}[\delta(X) \mid \theta] - g(\theta)
$$

Definition (不偏推定量)

推定量 $\delta\colon \mathcal{X} \to \mathbb{R}$ が以下を満たすとき、$g(\theta)$ の不偏推定量であるという。

$$
\mathbb{E}[\delta(X) \mid \theta] = g(\theta) \qquad \forall\theta \in \Theta.
$$

標本平均

$g(\theta) = \mathbb{E}[X \mid \theta]$ とすると、$\delta(X) = X$ は不偏推定量。独立に N 個サンプルする場合を考える。

$$p(\mathbf{x} \mid \theta) = \prod_{k=1}^{N} p(x_k \mid \theta).$$

$$\delta(\mathbf{X}) = \frac{1}{N} \sum_k X_k \qquad (\text{標本平均})$$

これも不偏推定量。分散は改善する。

$$\mathbb{V}\left[\frac{1}{N}\sum_k X_k \mid \theta\right] = \frac{1}{N^2}\mathbb{V}\left[\sum_k X_k \mid \theta\right] = \frac{1}{N}\mathbb{V}[X \mid \theta].$$

標本分散、不偏分散

$g(\theta) = \mathbb{V}[X \mid \theta]$ とする。

$$\delta(\mathbf{X}) = \frac{1}{N}\sum_k \left(X_k - \frac{1}{N}\sum_\ell X_\ell\right)^2 \qquad \text{(標本分散)}$$

$$\begin{aligned}
\mathbb{E}[\delta(\mathbf{X}) \mid \theta] &= \sum_k \mathbb{E}\left[\frac{1}{N}\left(X_k - \frac{1}{N}\sum_\ell X_\ell\right)^2 \mid \theta\right] \\
&= \frac{1}{N}\sum_k \mathbb{E}\left[\left(X_k - \mathbb{E}[X \mid \theta] + \mathbb{E}[X \mid \theta] - \frac{1}{N}\sum_\ell X_\ell\right)^2 \mid \theta\right] \\
&= \frac{1}{N}\sum_k \mathbb{E}\left[(X_k - \mathbb{E}[X \mid \theta])^2 \mid \theta\right] + \mathbb{E}\left[\left(\mathbb{E}[X \mid \theta] - \frac{1}{N}\sum_\ell X_\ell\right)^2 \mid \theta\right] \\
&\quad + \frac{2}{N}\sum_k \mathbb{E}\left[(X_k - \mathbb{E}[X \mid \theta])\left(\mathbb{E}[X \mid \theta] - \frac{1}{N}\sum_\ell X_\ell\right) \mid \theta\right] \\
&= \mathbb{V}[X \mid \theta] - \mathbb{E}\left[\left(\mathbb{E}[X \mid \theta] - \frac{1}{N}\sum_\ell X_\ell\right)^2 \mid \theta\right] \\
&= \mathbb{V}[X \mid \theta] - \mathbb{V}\left[\frac{1}{N}\sum_\ell X_\ell \mid \theta\right] = \frac{N-1}{N}\mathbb{V}[X \mid \theta].
\end{aligned}$$

$$\delta(\mathbf{X}) = \frac{1}{N-1}\sum_k \left(X_k - \frac{1}{N}\sum_\ell X_\ell\right)^2 \qquad \text{(不偏分散)}$$

不偏推定量は存在するとは限らない

パラメータ $\theta \in (0,1)$ について、$X \sim \mathrm{Ber}(\theta)$、$g(\theta) = 1/\theta$ とする。
$\delta\colon \{0,1\} \to \mathbb{R}$ を不偏推定量とすると、

$$\frac{1}{\theta} = \mathbb{E}\left[\delta(X) \mid \theta\right] = (1-\theta)\delta(0) + \theta\delta(1)$$

$\theta \to 0$ の極限で左辺は $+\infty$ に発散する。一方で右辺は $\delta(0)$ に収束する。よって δ をどのように定めても不偏推定量にはならない。

不偏推定量の自由度

Fact

$\delta_0 : \mathcal{X} \to \mathbb{R}$ を $g(\theta)$ の不偏推定量とする。任意の推定量 $\delta : \mathcal{X} \to \mathbb{R}$ について以下は同値

1. δ は $g(\theta)$ の不偏推定量
2. $\delta - \delta_0$ は 0 の不偏推定量

よって $h : \mathcal{X} \to \mathbb{R}$ を 0 の不偏推定量とすると、$g(\theta)$ の任意の不偏推定量は $\delta = \delta_0 + h$ と表せる。このとき、

$$\begin{aligned}
\mathbb{V}[\delta(X) \mid \theta] &= \mathbb{V}[\delta_0(X) + h(X) \mid \theta] \\
&= \mathbb{V}[\delta_0(X) \mid \theta] + \mathbb{V}[h(X) \mid \theta] + 2\mathrm{Cov}[\delta_0(X), h(X) \mid \theta] \\
&= \mathbb{V}[\delta_0(X) \mid \theta] + \mathbb{E}\left[h(X)^2 \mid \theta\right] + 2\mathbb{E}[\delta_0(X)h(X) \mid \theta]
\end{aligned}$$

よって $\mathbb{E}\left[h(X)^2 \mid \theta\right] + 2\mathbb{E}[\delta_0(X)h(X) \mid \theta]$ を最小化する h を見つければよい。

例: 不偏推定量の最適化 1/2

パラメータ $\theta \in (0,1)$ について、

$$\begin{aligned}\Pr(X = -1 \mid \theta) &= \theta \\ \Pr(X = x \mid \theta) &= \theta^x (1-\theta)^2 \qquad \forall x \in \mathbb{Z}_{\geq 0}\end{aligned}$$

とする (確率 θ で -1、確率 $1-\theta$ で幾何分布 $\mathrm{Geo}(\theta)$)。このとき、θ と $(1-\theta)^2$ の不偏推定量をそれぞれ

$$\begin{aligned}\delta_0(x) &= \mathbb{1}_{\{x=-1\}} \\ \delta_1(x) &= \mathbb{1}_{\{x=0\}}\end{aligned}$$

とする。また、$h : \mathbb{Z}_{\geq -1} \to \mathbb{R}$ を 0 の不偏推定量とすると、

$$\begin{aligned}& \theta h(-1) + \sum_{x \geq 0} h(x) \theta^x (1-\theta)^2 = 0 \qquad \forall \theta \in (0,1) \\ \iff & h(0) + \sum_{x \geq 1} (h(x) - 2h(x-1) + h(x-2)) \theta^x = 0 \qquad \forall \theta \in (0,1) \\ \iff & h(0) = 0, \qquad h(x) - 2h(x-1) + h(x-2) = 0 \qquad \forall x \in \mathbb{Z}_{\geq 1} \\ \iff & h(x) = xh(1) \qquad \forall x \in \mathbb{Z}_{\geq -1}\end{aligned}$$

例: 不偏推定量の最適化 2/2

$$
\begin{aligned}
&\mathbb{E}\left[h(X)^2 \mid \theta\right] + 2\mathbb{E}\left[\delta_0(X)h(X) \mid \theta\right] \\
&= \mathbb{E}\left[(h(1)X)^2 \mid \theta\right] + 2\mathbb{E}\left[\delta_0(X)h(1)X \mid \theta\right] \\
&= \mathbb{E}\left[X^2 \mid \theta\right] h(1)^2 + 2\mathbb{E}\left[\delta_0(X)X \mid \theta\right] h(1)
\end{aligned}
$$

よって $h(1) = -\frac{\mathbb{E}[\delta_0(X)X|\theta]}{\mathbb{E}[X^2|\theta]} = \frac{1-\theta}{2}$ のとき最小化される。最適な不偏推定量は θ 毎に違う。

同様に δ_1 について考えると $h(1) = -\frac{\mathbb{E}[\delta_1(X)X|\theta]}{\mathbb{E}[X^2|\theta]} = 0$ のとき最小化される。最適な不偏推定量は θ によらず $\delta_1(x) = \mathbb{1}_{\{x=0\}}$.

一様最小分散不偏推定量

Definition (一様最小分散不偏推定量 (uniform minimum variance unbiased (UMVU) estimator))

不偏推定量 $\delta\colon \mathcal{X} \to \mathbb{R}$ が以下の条件を満たすとき**一様最小分散不偏推定量**という。

1. (不偏推定量である) $\mathbb{E}\left[\delta(X) \mid \theta\right] = g(\theta)$ for all $\theta \in \Theta$.
2. (分散が最小である) **任意の不偏推定量** $\delta'\colon \mathcal{X} \to \mathbb{R}$ と $\theta \in \Theta$ について $\mathbb{V}\left[\delta(X) \mid \theta\right] \leq \mathbb{V}\left[\delta'(X) \mid \theta\right]$ が成り立つ。

一様最小分散不偏推定量

Theorem

$$\Delta := \{\delta : \mathcal{X} \to \mathbb{R} \mid \mathbb{E}\left[\delta(X)^2 \mid \theta\right] < \infty \quad \forall\theta \in \Theta\}$$
$$\mathcal{U} := \{h : \mathcal{X} \to \mathbb{R} \mid \mathbb{E}\left[h(X) \mid \theta\right] = 0 \quad \forall\theta \in \Theta\} \cap \Delta$$

とする。$\delta \in \Delta$ について以下は同値

1. δ は $\mathbb{E}\left[\delta(X) \mid \theta\right]$ の *UMVU* 推定量。
2. 任意の $h \in \mathcal{U}$ と $\theta \in \Theta$ について $\mathbb{E}\left[\delta(X)h(X) \mid \theta\right] = 0$。

Proof.

$1 \implies 2$: δ は $\mathbb{E}\left[\delta(X) \mid \theta\right]$ の UMVU 推定量と仮定する。任意の $h \in \mathcal{U}$ と $\theta \in \Theta$ と $a \in \mathbb{R}$ について

$$\begin{aligned}\mathbb{V}\left[\delta(X) \mid \theta\right] &\leq \mathbb{V}\left[\delta(X) + ah(X) \mid \theta\right] \\ &= \mathbb{V}\left[\delta(X) \mid \theta\right] + \mathbb{E}\left[h(X)^2 \mid \theta\right] a^2 + 2\mathbb{E}\left[\delta(X)h(X) \mid \theta\right] a\end{aligned}$$

が成り立つので $(h(x) = 0$ のときとそうでないときで場合分けすると$)$、$\mathbb{E}\left[\delta(X)h(X) \mid \theta\right] = 0$。 □

一様最小分散不偏推定量

Theorem

$$\Delta := \{\delta : \mathcal{X} \to \mathbb{R} \mid \mathbb{E}\left[\delta(X)^2 \mid \theta\right] < \infty \quad \forall \theta \in \Theta\}$$
$$\mathcal{U} := \{h : \mathcal{X} \to \mathbb{R} \mid \mathbb{E}\left[h(X) \mid \theta\right] = 0 \quad \forall \theta \in \Theta\} \cap \Delta$$

とする。$\delta \in \Delta$ について以下は同値

1. δ は $\mathbb{E}\left[\delta(X) \mid \theta\right]$ の *UMVU* 推定量。
2. 任意の $h \in \mathcal{U}$ と $\theta \in \Theta$ について $\mathbb{E}\left[\delta(X)h(X) \mid \theta\right] = 0$。

Proof.

$2 \implies 1$: 任意の $h \in \mathcal{U}$ と $\theta \in \Theta$ について $\mathbb{E}\left[\delta(X)h(X) \mid \theta\right] = 0$ と仮定する。$\delta' : \mathcal{X} \to \mathbb{R}$ を $\mathbb{E}\left[\delta(X) \mid \theta\right]$ の不偏推定量とする。$\mathbb{V}\left[\delta'(X) \mid \theta\right] = \infty$ の場合は $\mathbb{V}\left[\delta(X) \mid \theta\right] \le \mathbb{V}\left[\delta'(X) \mid \theta\right]$ である。$\delta' \in \Delta$ の場合は $h := \delta' - \delta \in \mathcal{U}$。このとき任意の $\theta \in \Theta$ について

$$\begin{aligned}
\mathbb{V}\left[\delta'(X) \mid \theta\right] &= \mathbb{V}\left[\delta(X) + h(X) \mid \theta\right] \\
&= \mathbb{V}\left[\delta(X)\right] + \mathbb{E}\left[h(X)^2 \mid \theta\right] + 2\mathbb{E}\left[\delta(X)h(X) \mid \theta\right] \\
&= \mathbb{V}\left[\delta(X)\right] + \mathbb{E}\left[h(X)^2 \mid \theta\right] \ge \mathbb{V}\left[\delta(X)\right]
\end{aligned}$$

推定量の平均

Lemma

$\delta_1, \delta_2, \dots, \delta_n : \mathcal{X} \to \Theta$ が $g(\theta)$ の不偏推定量のとき、

$$\delta^* := \frac{1}{n} \sum_{k=1}^{n} \delta_k$$

は $g(\theta)$ の不偏推定量で

$$\mathbb{V}\left[\delta^*(X) \mid \theta\right] \leq \frac{1}{n} \sum_{k=1}^{n} \mathbb{V}\left[\delta_k(X) \mid \theta\right] \qquad \forall \theta \in \Theta.$$

Proof.

δ^* が不偏推定量であることは期待値の線形性から明らか。

$$\begin{aligned}
\mathbb{V}\left[\delta^*(X) \mid \theta\right] &= \mathbb{E}\left[\left(\frac{1}{n} \sum_{k=1}^{n} \delta_k(X) - g(\theta)\right)^2 \mid \theta\right] \\
&\leq \frac{1}{n} \sum_{k=1}^{n} \mathbb{E}\left[(\delta_k(X) - g(\theta))^2 \mid \theta\right] \quad \text{(イェンセンの不等式)} \quad \square
\end{aligned}$$

対称な不偏推定量

$$p(\mathbf{x} \mid \theta) = \prod_{k=1}^{N} p(x_k \mid \theta).$$

Lemma (対称な不偏推定量)

任意の不偏推定量 $\delta\colon \mathcal{X}^N \to \Theta$ について、ある $(x_1, \ldots, x_n$ について$)$ 対称な不偏推定量 δ^* が存在して

$$\mathbb{V}\left[\delta^*(\mathbf{X}) \mid \theta\right] \leq \mathbb{V}\left[\delta(\mathbf{X}) \mid \theta\right] \qquad \forall \theta \in \Theta.$$

Proof.

$$\delta^*(x_1, \ldots, x_N) := \frac{1}{N!} \sum_{\sigma \in S_N} \delta\left(x_{\sigma(1)}, \ldots, x_{\sigma(N)}\right)$$

と定義すればよい。 □

コーシー–シュワルツの不等式

Lemma

任意の確率変数 X と Y について

$$|\mathbb{E}[XY]| \leq \sqrt{\mathbb{E}[X^2]\,\mathbb{E}[Y^2]}.$$

Proof.

任意の $a \in \mathbb{R}$ について

$$\begin{aligned} 0 &\leq \mathbb{E}\left[(X+aY)^2\right] \\ &\leq \mathbb{E}\left[X^2\right] + \mathbb{E}\left[Y^2\right]a^2 + 2\mathbb{E}[XY]\,a \end{aligned}$$

$\mathbb{E}\left[Y^2\right] = 0$ のとき、$\Pr(Y = 0) = 1$ で不等式は成り立つ。$\mathbb{E}\left[Y^2\right] > 0$ とする。上の式に $a = -\frac{\mathbb{E}[XY]}{\mathbb{E}[Y^2]}$ を代入すると、

$$0 \leq \mathbb{E}\left[X^2\right] - \frac{\mathbb{E}[XY]^2}{\mathbb{E}[Y^2]}$$

よって不等式は得られた。 □

共分散の不等式

Corollary

任意の確率変数 X と Y について

$$|\mathrm{Cov}[X, Y]| \leq \sqrt{\mathbb{V}[X]\,\mathbb{V}[Y]}.$$

Proof.

$A = X - \mathbb{E}[X]$, $B = Y - \mathbb{E}[Y]$ について、コーシー–シュワルツの不等式より

$$|\mathbb{E}[AB]| \leq \sqrt{\mathbb{E}[A^2]\,\mathbb{E}[B^2]}$$

$$\iff |\mathrm{Cov}[X, Y]| \leq \sqrt{\mathbb{V}[X]\,\mathbb{V}[Y]}.$$

□

不偏推定量の分散の下界

不偏推定量 $\delta\colon \mathcal{X} \to \mathbb{R}$ の分散の下界を導出したい。

任意の $\psi\colon \mathcal{X} \times \Theta \to \mathbb{R}$ について

$$\mathbb{V}[\delta(X) \mid \theta] \geq \frac{\mathrm{Cov}[\delta(X), \psi(X, \theta) \mid \theta]^2}{\mathbb{V}[\psi(X, \theta) \mid \theta]}.$$

ここで $p(x \mid \theta)$ が θ で微分可能なとき

$$\psi(x, \theta) = \frac{\partial \log p(x \mid \theta)}{\partial \theta} = \frac{1}{p(x \mid \theta)} \frac{\partial p(x \mid \theta)}{\partial \theta}$$

とおく (離散の場合は $p(x \mid \theta)$ は確率質量関数としての尤度関数に置き換える) と、

$$\begin{aligned} \mathbb{E}[\psi(X, \theta) \mid \theta] &= \int \frac{\partial p(x \mid \theta)}{\partial \theta} \mathrm{d}x \\ &= \frac{\partial}{\partial \theta} \int p(x \mid \theta) \mathrm{d}x = 0. \qquad \text{(微分と積分の交換)} \end{aligned}$$

フィッシャー情報量

Definition (フィッシャー情報量)

$$I(\theta) := \mathbb{E}\left[\left(\frac{\partial \log p(X \mid \theta)}{\partial\theta}\right)^2 \mid \theta\right].$$

$$\begin{aligned}\mathbb{V}[\delta(X) \mid \theta] &\geq \frac{\mathrm{Cov}[\delta(X), \psi(X, \theta) \mid \theta]^2}{\mathbb{V}[\psi(X, \theta) \mid \theta]} \\ &= \frac{\mathbb{E}[\delta(X)\psi(X, \theta) \mid \theta]^2}{I(\theta)}.\end{aligned}$$

$$\begin{aligned}\mathbb{E}[\delta(X)\psi(X, \theta) \mid \theta] &= \int p(x \mid \theta)\delta(x)\frac{1}{p(x \mid \theta)}\frac{\partial p(x \mid \theta)}{\partial\theta}\mathrm{d}x \\ &= \frac{\partial}{\partial\theta}\int p(x \mid \theta)\delta(x)\mathrm{d}x \qquad \text{(微分と積分の交換)} \\ &= \frac{\partial}{\partial\theta}g(\theta) = g'(\theta).\end{aligned}$$

クラメール–ラオの不等式

Theorem (クラメール–ラオの不等式)

$\delta\colon \mathcal{X} \to \mathbb{R}$ を $\mathrm{g}(\theta)$ の任意の不偏推定量とするとき以下が成り立つ。

$$\mathbb{V}\left[\delta(X) \mid \theta\right] \geq \frac{\mathrm{g}'(\theta)^2}{I(\theta)}.$$

課題

パラメータ $\theta \in (0, 1)$ について、$X_1, \ldots, X_N$ が独立なベルヌーイ分布 $\mathrm{Ber}(\theta)$ に従うものとする。つまり、以下が成り立つ。

$$\Pr(X = x_1, \ldots, X_N = x_N \mid \theta) = \theta^{x_1+\cdots+x_N}(1-\theta)^{N-(x_1+\cdots+x_N)}.$$

$\mathbf{X}$ から $g(\theta)$ を推定する問題を考える。$T(\mathbf{x}) = \frac{1}{N}\sum_k x_k$ とする。以下の問に答えよ。

1. $\mathbf{X}$ のフィッシャー情報量 $I(\theta)$ をもとめよ。
2. $g(\theta) = \mathbb{E}[X \mid \theta] = \theta$ の不偏推定量 $\delta(\mathbf{x}) = T(\mathbf{x})$ の分散とクラメール–ラオの下界を導出せよ。
3. $g(\theta) = \mathbb{V}[X \mid \theta] = \theta(1-\theta)$ の不偏推定量である不偏分散 $\delta(\mathbf{x}) = \frac{1}{N-1}\sum_k (x_k - T(\mathbf{x}))^2$ を $T(\mathbf{x})$ の関数として表せ。また、$g(\theta)$ の不偏推定量の分散に対するクラメール–ラオの下界を導出せよ。